NTP
National Toxicology Program

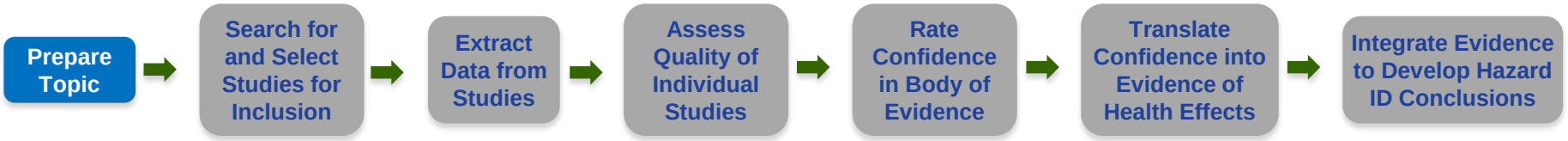
Draft OHAT Approach
Part 1
Preparing the Topic Through Assessing the Quality of Individual Studies

Abee Boyles, Ph.D.
Division of the National Toxicology Program,
Office of Health Assessment and Translation (OHAT)

Web-Based Informational Meeting
April 23, 2013 12:00 - 4:00PM EDT

Step 1. Prepare the Topic

- **Scope and focus the topic** to answer specific questions
  - Consult with appropriate experts to focus objective and questions
- **Develop draft protocol** through iterative process to detail project-specific procedures used throughout the evaluation
  - Literature search, selection of relevant studies
  - Determination of primary, secondary, and grouping of outcomes
  - Data extraction, assessment of risk of bias, evaluation of confidence
- **Obtain input on draft protocol** from experts (technical advisors, BSC) and solicit public comment
  - Draft protocols illustrate the application of this framework
  - Protocols available at: http://ntp.niehs.nih.gov/go/38673
- **Document** project-specific procedures before proceeding with the evaluation



Importance of the Protocol

- **The protocol contains enough details so that the process and the procedures could be reconstructed**
  - For example
    - The literature search strategy is presented in enough detail so that it could be replicated
    - State which outcomes are considered primary and secondary
    - Criteria for assessing individual study quality are established stating what defines high and low risk of bias for each question on study design or performance
- **Revisions to the protocol**
  - It is recognized that valid reasons for modifying a protocol may occur during an evaluation
  - Revisions are permitted and they are documented and justified

Primary and Secondary Outcomes

Example – PFOA/PFOS Exposure and Immunotoxicity

Humans	Animals*	*In vitro* Assays
Primary outcomes	*Primary outcomes*	*Primary outcomes*
Immune-related diseases and measures of immune function • *Immunosuppression* • *Sensitization and allergic response* • *Autoimmunity*	**Disease resistance assay or measures of immune function** • *Disease resistance assays* • *Immune function assays following in vivo exposure to the test substance*	***Immune function assays following in vitro exposure to the test substance***
Secondary outcomes	*Secondary outcomes*	*Secondary outcomes*
Immunostimulation** ***Observational immune endpoints***	***Observational immune endpoints following in vivo exposure to test substance***	***Observational immune endpoints following in vitro exposure to the test substance***

Primary and Secondary Outcomes

Example – PFOA/PFOS Exposure and Immunotoxicity

More detail and examples provided in the protocol

Humans	Animals*	*In vitro* Assays
Primary outcomes	*Primary outcomes*	*Primary outcomes*
Immune-related diseases and measures of immune function *Immunosuppression* (e.g., otitis, infections, or decreased vaccine antibody response); *Sensitization and allergic response* (e.g., atopic dermatitis or asthma); *Autoimmunity* (e.g., thyroiditis or systemic lupus erythematosus)	Disease resistance assay or measures of immune function *Disease resistance assays* (e.g., host resistance to influenza A or trichinella, changes in incidence or progression in animal models of autoimmune disease) *Immune function assays following <u>in vivo exposure</u> to the test substance* (e.g., antibody response [T-cell dependent IgM antibody response (TDAR)], natural killer cell [NK] activity, delayed-type hypersensitivity [DTH] response, phagocytosis by monocytes, local lymph-node assay [LLNA])	*Immune function assays following <u>in vitro exposure</u> to the test substance* (e.g., natural killer cell [NK] activity, phagocytosis or bacterial killing by monocytes, proliferation following anti-CD3 antibody stimulation of spleen cells or lymphocytes)
Secondary outcomes	*Secondary outcomes*	*Secondary outcomes*
*Immunostimulation*** (e.g., unintended stimulation of humoral immune function) *Observational immune endpoints* (e.g., lymphocyte counts, lymphocyte proliferation, cytokine levels, serum antibody levels, or serum autoantibody levels)	*Observational immune endpoints* (e.g., lymphoid organ weight, lymphocyte counts or subpopulations, lymphocyte proliferation, cytokine production, serum antibody levels, serum or tissue autoantibody levels, or histopathological changes in immune organs)	*Observational immune endpoints following <u>in vitro exposure</u> to the test substance* (e.g., general mitogen-stimulated lymphocyte proliferation, cytokine production)

Primary and Secondary Outcomes
Example – BPA Exposure and Obesity

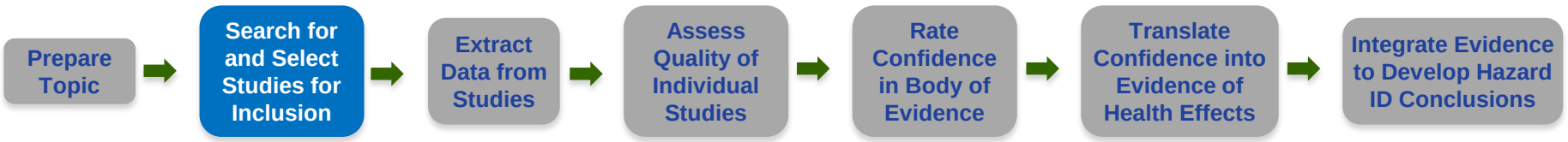
More detail and examples provided in the protocol

Humans	Animals	Supporting Evidence
Primary outcomes	*Primary outcomes*	*Phenotypic or "apical" outcomes*
overweight, obesity measures, or measures of adiposity (e.g., BMI, waist circumference, fat composition, skin-fold thickness)	adiposity (e.g., fat mass, percent fat)	e.g., adipogenic endpoints such as adipocyte number, adipocyte differentiation, or adipocyte lipid accumulation
Secondary outcomes	*Secondary outcomes*	*Pathway and cellular endpoints*
adipokines, ghrelin, leptin, adiponectin, resistin, feeding behavior	adipokines, ghrelin, leptin, adiponectin, resistin, feeding behavior, body weight	e.g., *ex vivo*, cellular, genomic, or mode of action outcomes reported in eligible animal or human studies; cellular, genomic, or mode of action outcomes reported in *in vitro* studies of adipocytes; interactions with key receptors involved in regulating adipogenesis, e.g., peroxisome proliferator-activated receptors (PPAR), *retinoid X receptor (RXR), liver X receptor (LXR),* or glucocorticoid receptor (GR) in any *in vitro* model or high throughput screening system

Step 2: Search for and Select Studies for Inclusion

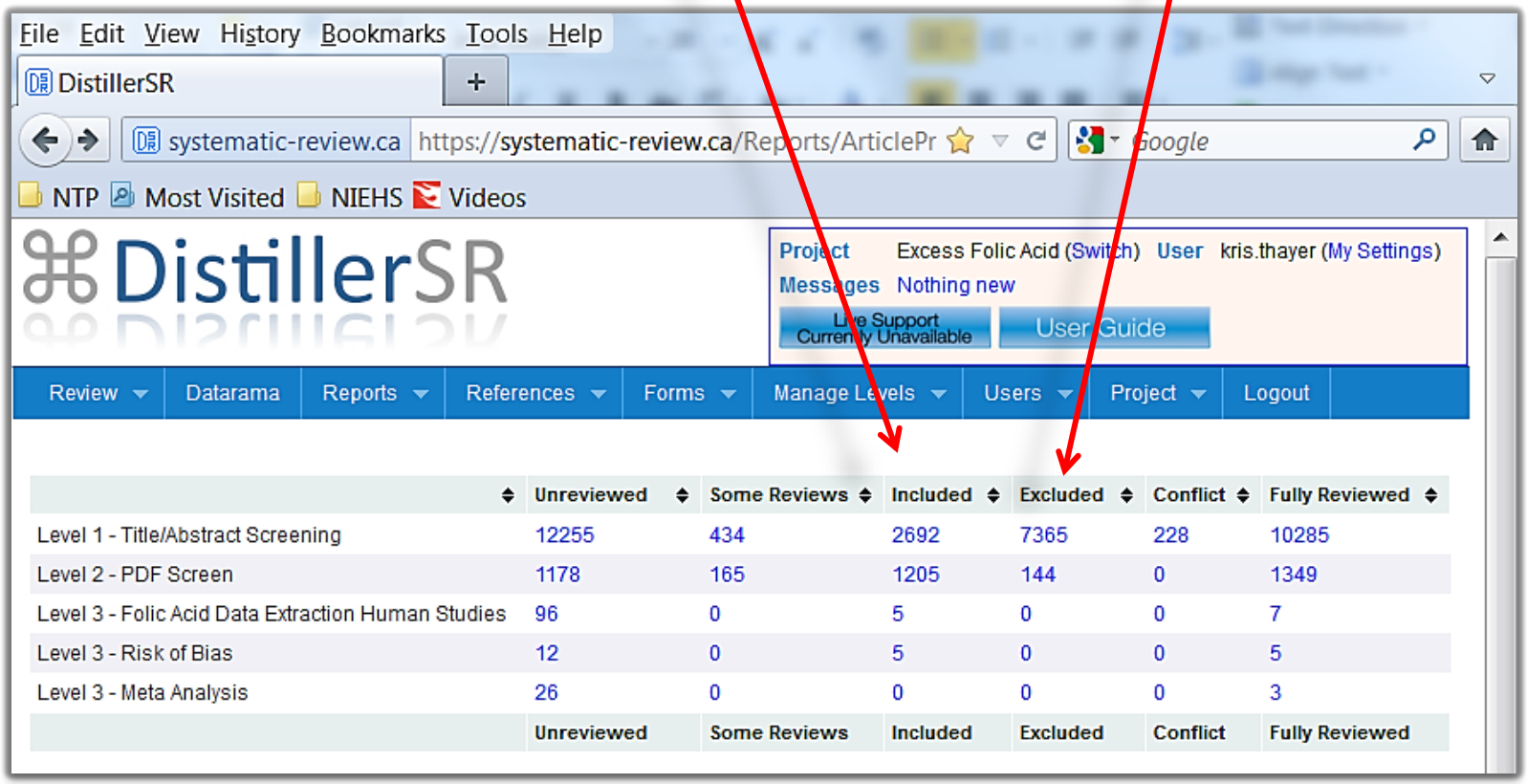
- **Perform comprehensive literature search**
  - Follow search strategy specified in protocol
    - Including search dates, frequency of updates, etc.
    - Use of unpublished studies – e.g., peer review of critical studies

- **Screen studies for inclusion**
  - Two reviewers evaluate each study at the title/abstract level
  - Follow procedures defined in protocol to
    - Select relevant studies based on pre-defined criteria
    - Resolve conflicts between reviewers
    - Document reasons for exclusion
    - Complete full-text review



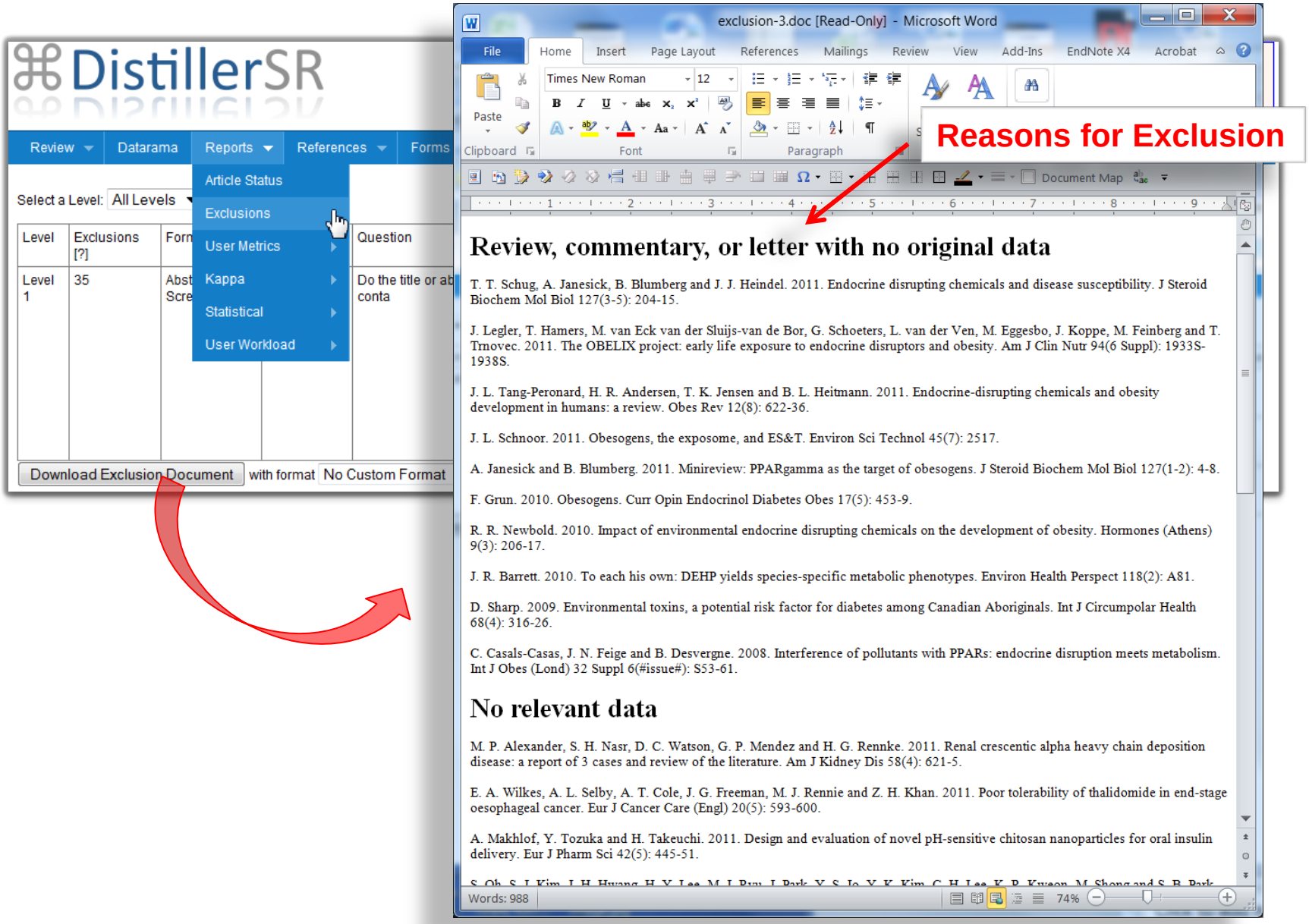
Web-based Reference Management

- DistillerSR® Systematic Review Software
- Project management and workflow
- Tracks which studies were included or why excluded

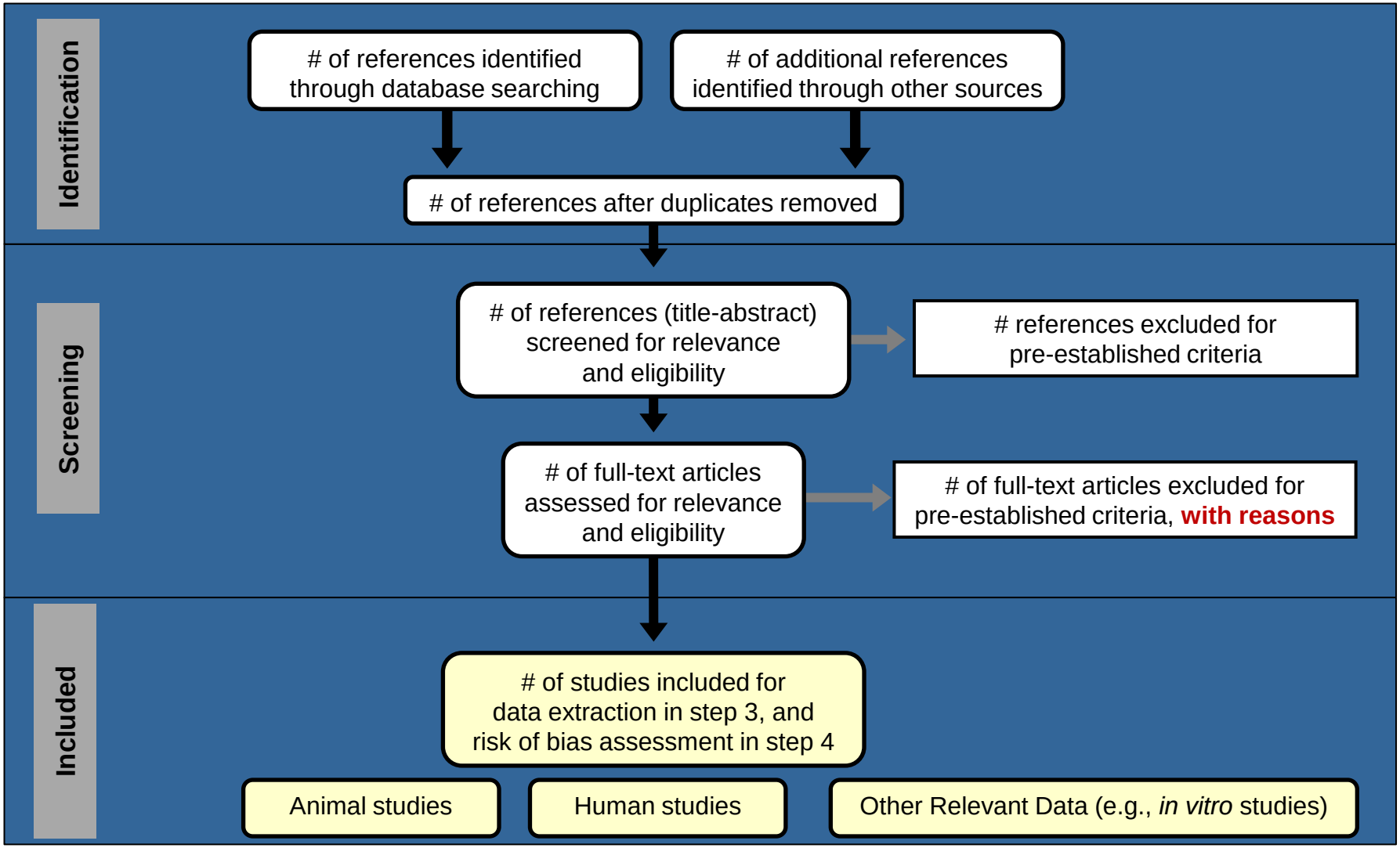


	Unreviewed	Some Reviews	Included	Excluded	Conflict	Fully Reviewed
Level 1 - Title/Abstract Screening	12255	434	2692	7365	228	10285
Level 2 - PDF Screen	1178	165	1205	144	0	1349
Level 3 - Folic Acid Data Extraction Human Studies	96	0	5	0	0	7
Level 3 - Risk of Bias	12	0	5	0	0	5
Level 3 - Meta Analysis	26	0	0	0	0	3
	Unreviewed	Some Reviews	Included	Excluded	Conflict	Fully Reviewed

Exclusion Report



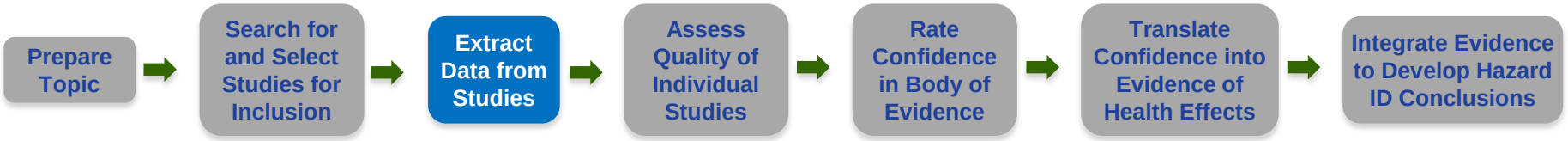
From Literature Search to Study Selection



Adapted from Moher D *et al.* 2009. Preferred reporting items for systematic reviews and meta-analyses: The PRISMA statement. *Journal of Clinical Epidemiology 62(10): 1006-1012.*

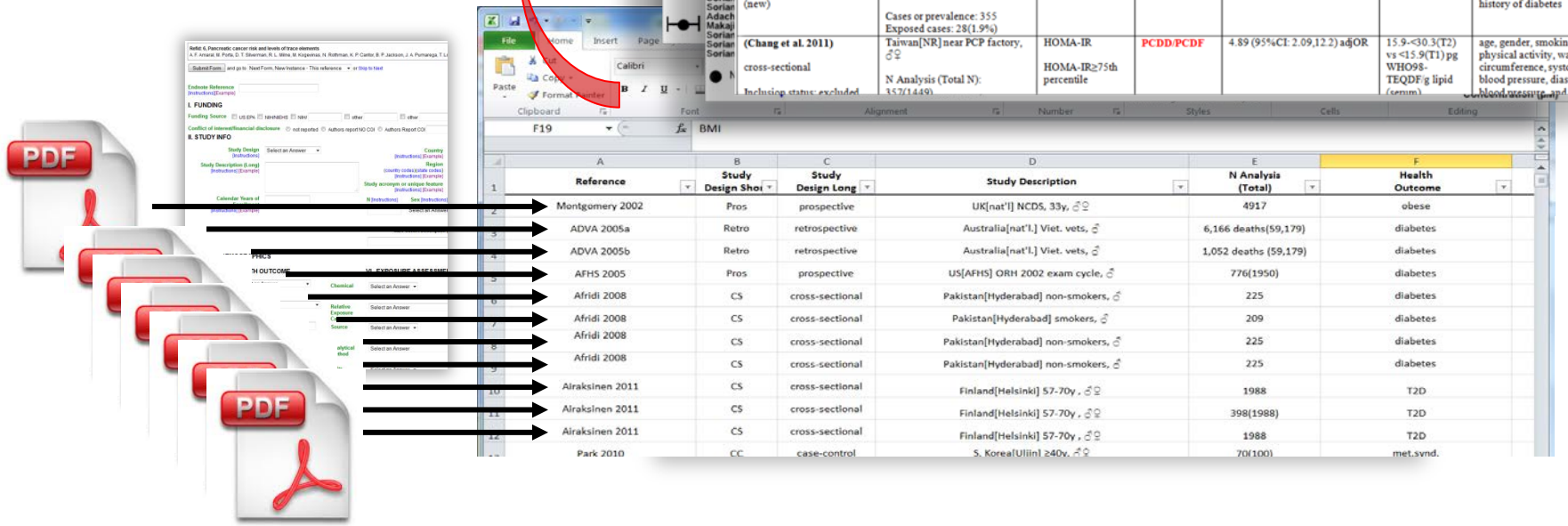
Step 3: Extract Data from Studies

- **Extract data**
  - Individual study information collected systematically
  - There are separate template data extraction forms for human, animal, and *in vitro* studies
  - Follow procedures defined in protocol for
    - Data extraction by a member of the evaluation team
    - Data extraction forms would be customized for each evaluation
    - Quality assurance of data



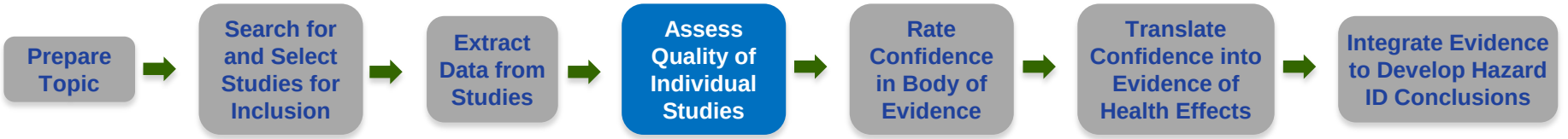
Data Extraction Applications

- Data extraction is transparent and consistent
- Files can be disseminated for data mining
- Variety of outputs without re-entering data
  - Figures (Meta Data Viewer)
  - Appendix tables (Mail Merge)



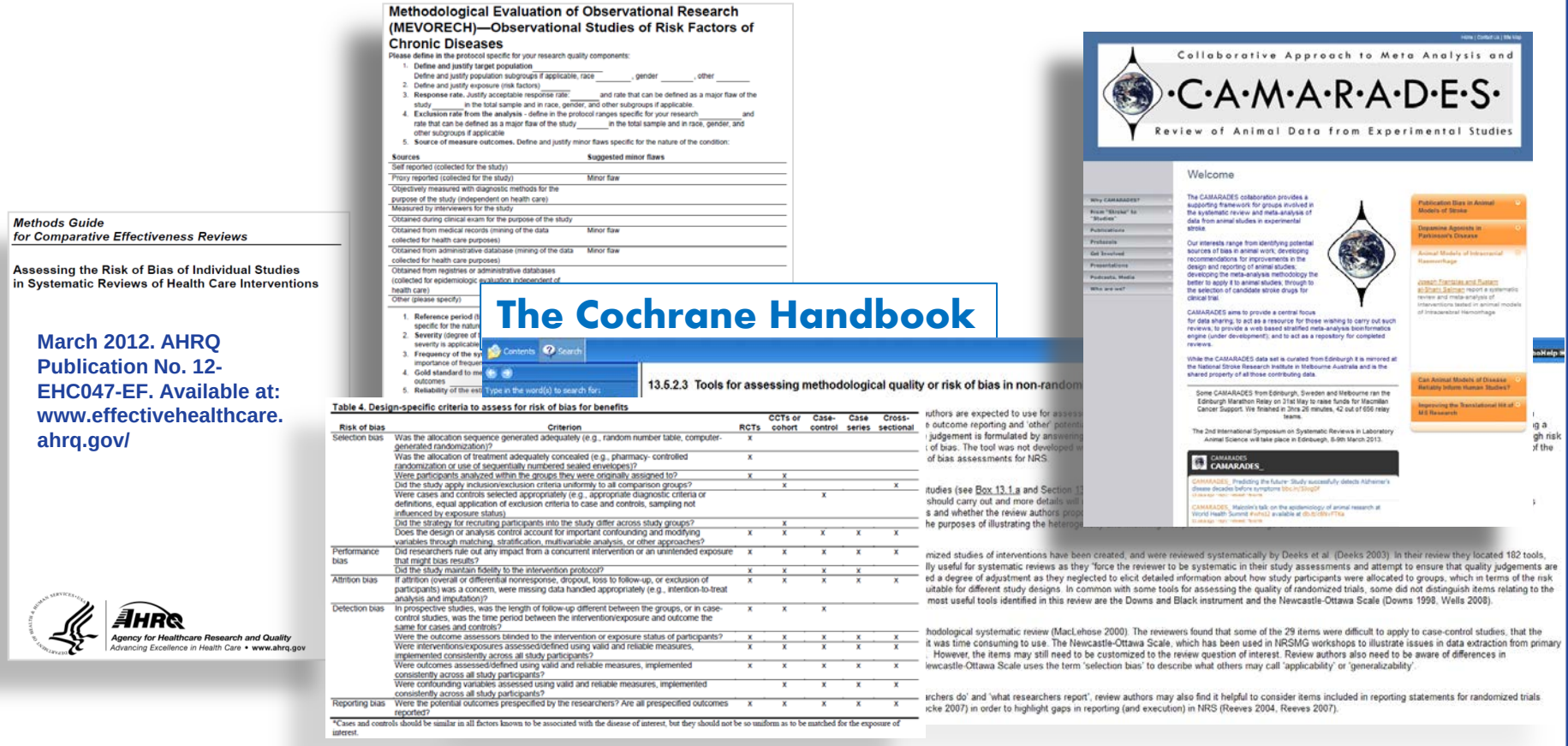
Step 4: Assess the Quality of Individual Studies

- **Study quality or risk of bias**
  - How credible are the study findings?
- **State of the art for assessing risk of bias**
  - Single summary scores for “study quality” are strongly discouraged
  - Reporting quality checklists are of limited utility (mix bias and reporting)
- **Existing methods**
  - Established risk of bias tools for randomized controlled trials
  - No existing consensus on how to assess risk of bias for
    - Observational human studies, or
    - Animal studies
    - *In vitro* studies



Adaptation of Existing Study Quality Methods

- Although there are a variety of risk of bias methods for human studies, animal tools are generally reporting quality checklists (e.g., ToxRTool)
- The recent AHRQ method guide* was particularly useful as a model because it covers RCTs and a range of human observational studies



* Assessing the Risk of Bias of Individual Studies in Systematic Reviews of Health Care Interventions (AHRQ, Viswanathan, 2012)

Adaptation of Existing Study Quality Methods

- Although there are a variety of risk of bias methods for human studies, animal tools are generally reporting quality checklists (e.g., ToxRTool)
- The recent AHRQ method guide* was particularly useful as a model because it covers RCTs and a range of human observational studies

Consideration of 5 traditional risk of bias domains

Study design determines which questions apply

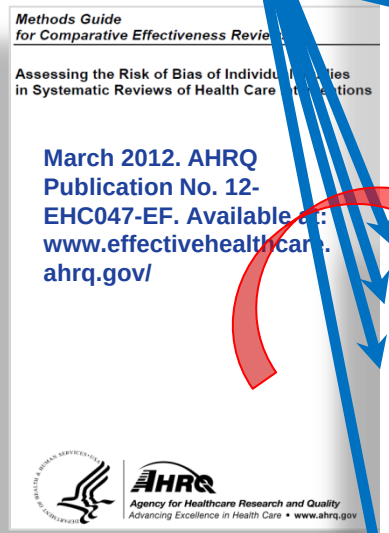


Table 4. Design-specific criteria to assess for risk of bias for benefits

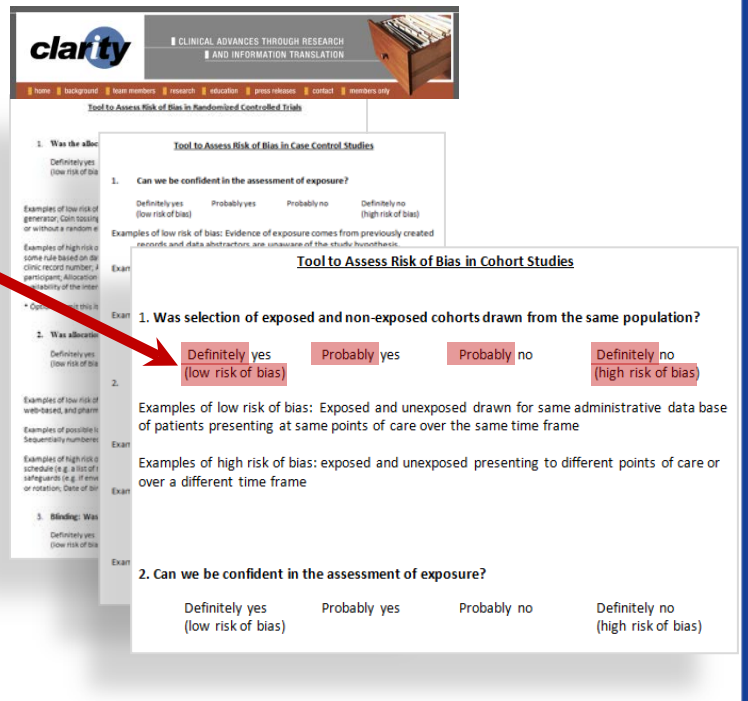
Risk of bias	Criterion	RCTs	CCTs or cohort	Case-control	Case series	Cross-sectional
Selection bias	Was the allocation sequence generated adequately (e.g., random number table, computer-generated randomization)?	x				
	Was the allocation of treatment adequately concealed (e.g., pharmacy- controlled randomization or use of sequentially numbered sealed envelopes)?	x				
	Were participants analyzed within the groups they were originally assigned to?	x	x			
	Did the study apply inclusion/exclusion criteria uniformly to all comparison groups?		x			x
	Were cases and controls selected appropriately (e.g., appropriate diagnostic criteria or definitions, equal application of exclusion criteria to case and controls, sampling not influenced by exposure status)			x		
	Did the strategy for recruiting participants into the study differ across study groups?		x			
	Does the design or analysis control account for important confounding and modifying variables through matching, stratification, multivariable analysis, or other approaches?	x	x	x	x	x
Performance bias	Did researchers rule out any impact from a concurrent intervention or an unintended exposure that might bias results?	x	x	x	x	x
	Did the study maintain fidelity to the intervention protocol?	x	x	x	x	
Attrition bias	If attrition (overall or differential nonresponse, dropout, loss to follow-up, or exclusion of participants) was a concern, were missing data handled appropriately (e.g., intention-to-treat analysis and imputation)?	x	x	x	x	x
Detection bias	In prospective studies, was the length of follow-up different between the groups, or in case-control studies, was the time period between the intervention/exposure and outcome the same for cases and controls?	x	x	x		
	Were the outcome assessors blinded to the intervention or exposure status of participants?	x	x	x	x	x
	Were interventions/exposures assessed/defined using valid and reliable measures, implemented consistently across all study participants?	x	x	x	x	x
	Were outcomes assessed/defined using valid and reliable measures, implemented consistently across all study participants?	x	x	x	x	x
	Were confounding variables assessed using valid and reliable measures, implemented consistently across all study participants?		x	x	x	x
Reporting bias	Were the potential outcomes prespecified by the researchers? Are all prespecified outcomes reported?	x	x	x	x	x

*Cases and controls should be similar in all factors known to be associated with the disease of interest, but they should not be so uniform as to be matched for the exposure of interest.

* Assessing the Risk of Bias of Individual Studies in Systematic Reviews of Health Care Interventions (AHRQ, Viswanathan, 2012)

Adaptation of Existing Study Quality Methods

- Although there are a variety of risk of bias methods for human studies, animal tools are generally reporting quality checklists (e.g., ToxRTool)
- The recent AHRQ method guide* was particularly useful as a model because it covers RCTs and a range of human observational studies
- The clarity group scale for answering risk of bias questions was also useful (definitely low, probably low, probably high, to definitely high)



* Assessing the Risk of Bias of Individual Studies in Systematic Reviews of Health Care Interventions (AHRQ, Viswanathan, 2012)

Step 4: Assess the Quality of Individual Studies

- **Study quality or risk of bias**
  - Judge whether the design and conduct of individual studies compromise credibility in the link between exposure and outcome
  - Evaluation is endpoint/outcome specific
  - Use predefined set of questions adapted from AHRQ to address **both human studies and animal toxicology studies**

- Study design determines which questions are applicable
- Answers equate to risk of bias rating for each question/criteria

++	**Definitely Low risk of bias**
+	**Probably Low risk of bias**
–	**Probably High risk of bias**
––	**Definitely High risk of bias**

Prepare Topic → Search for and Select Studies for Inclusion → Extract Data from Studies → **Assess Quality of Individual Studies** → Rate Confidence in Body of Evidence → Translate Confidence into Evidence of Health Effects → Integrate Evidence to Develop Hazard ID Conclusions

General Risk of Bias Answer Format

++ Definitely Low:

- Direct evidence of low risk of bias practices
  (Protocol includes specific examples of relevant low risk of bias practices)

+ Probably Low:

- Indirect evidence of low risk of bias practices
- OR deviations from low bias practices would not appreciably bias results

- Probably High:

- Indirect evidence of high risk of bias practices
- OR there is insufficient information provided

-- Definitely High:

- Direct evidence of high risk of bias practices

Selective Reporting Bias
Example Question – Appendix 2 of Protocols

Were all measured outcomes reported?

- **++ Definitely Low:**
  - There is direct evidence that all of the study's measured outcomes (primary and secondary) outlined in the protocol, methods, abstract, and/or introduction (that are relevant for the evaluation) have been reported. This would include outcomes reported with sufficient detail to be included in meta-analysis or fully tabulated during data extraction.

Explicit guidance is provided for each answer to determine the risk of bias rating

Selective Reporting Bias

Example – Appendix 2 of Protocols (continued)

- **Were all measured outcomes reported?**

++ Definitely Low:

- There is direct evidence that all of the study's measured outcomes (primary and secondary) outlined in the protocol, methods, abstract, and/or introduction (that are relevant for the evaluation) have been reported. This would include outcomes reported with sufficient detail to be included in meta-analysis or fully tabulated during data extraction.

+ Probably Low:

- There is indirect evidence that all of the study's measured outcomes (primary and secondary) outlined in the protocol, methods, abstract, and/or introduction (that are relevant for the evaluation) have been reported
- **OR** analyses that had not been planned at the outset of the study (i.e., retrospective unplanned subgroup analyses) are clearly indicated as such and it is deemed that the omitted analyses were not appropriate and selective reporting would not appreciably bias results. This would include outcomes reported with insufficient detail such as only reporting that results were statistically significant (or not).

Selective Reporting Bias

Example – Appendix 2 of Protocols (continued)

- **Were all measured outcomes reported?**

– **Probably High:**

  – There is indirect evidence that all of the study's measured outcomes (primary and secondary) outlined in the protocol, methods, abstract, and/or introduction (that are relevant for the evaluation) have not been reported

  – **OR** there is insufficient information provided about selective outcome reporting.

-- **Definitely High:**

  – There is direct evidence that all of the study's measured outcomes (primary and secondary) outlined in the protocol, methods, abstract, and/or introduction (that are relevant for the evaluation) have not been reported. In addition to not reporting outcomes, this would include reporting outcomes based on composite score without individual outcome components or outcomes reported using measurements, analysis methods or subsets of the data (e.g., subscales) that were not pre-specified or reporting outcomes not pre-specified (unless clear justification for their reporting is provided, such as an unexpected effect).

Risk of Bias Rating Individual Animal Studies

++ Definitely Low risk of bias
+ Probably Low risk of bias
- Probably High risk of bias
-- Definitely High risk of bias
Not applicable due to study design

Draft OHAT Risk of Bias Questions	Andy .et al., 2010	Bucher et al., 1999	Wolfe et al., 2000	Boyles et al., 2011	Thayer et al., 2008
Was administered dose or exposure adequately randomized?	+				
Was allocation to the study groups adequately concealed?	-				

Risk of Bias Rating Individual Animal Studies

++ Definitely Low risk of bias
\+ Probably Low risk of bias
\- Probably High risk of bias
-- Definitely High risk of bias
○ Not applicable due to study design

Draft OHAT Risk of Bias Questions	Andy .et al., 2010	Bucher et al., 1999	Wolfe et al., 2000	Boyles et al., 2011	Thayer et al., 2008
Selection Bias					
Was administered dose or exposure level adequately randomized?	+	++	+	+	-
Was allocation to study groups adequately concealed?	-	++	+	--	+
Were the comparison groups appropriate?	○	○	○	○	○
Confounding Bias					
Did the study design or analysis account for important confounding and modifying variables?	-	++	-	--	+
Did researchers adjust or control for other exposures that are anticipated to bias results?	-	++	-	-	--
Performance Bias					
Were experimental conditions identical across study groups?	--	++	-	-	+
Did deviations from the study protocol impact the results?	-	++	+	-	-
Were the research personnel and human subjects blinded to the study group during the study?	-	++	+	+	+
Attrition / Exclusion Bias					
Were outcome data incomplete due to attrition or exclusion from analysis?	--	++	-	+	+
Information / Detection Bias					
Were outcome assessors blinded to study group or exposure group?	+	++	+	+	+
Were confounding variables assessed consistently across groups using valid and reliable measures?	--	++	+	++	++
Can we be confident in the exposure characterization?	-	++	--	-	+
Can we be confident in the outcome assessment?	-	++	-	+	-
Selective Reporting Bias					
Were all measured outcomes reported?	+	++	+	-	+

Using Risk of Bias Data

- **Ability to categorize or "tier" studies based on risk of bias**
  - Can clearly show risk of bias for individual factors or "tier" by all factors
  - Can define "key" risk of bias questions on a project-specific basis
- **Enhance transparency** – risk of bias released as part of evaluation
- **Can stratify or restrict confidence rating conclusions**
  - Stratified analysis with high risk of bias studies included to assess impact
  - Use studies with lower risk of bias (1st and 2nd tier)

Example of Tiers for Risk of Bias

Category	Guidance	Key Criteria		
1st tier	"definitely low" or "probably low" risk of bias for key criteria **AND** "definitely low" or "probably low" risk of bias for ≥50% of other (++ + + + + + – –)	+	++	+
2nd tier	Study does not meet criteria for 1st or 2nd tier			
3rd tier	"definitely high" or "probably high" risk of bias for key criteria **AND** "definitely high" or "probably high" risk of bias for ≥50% of other (++ + + – – – – ––)	–	–	––

- **Observational Studies (most human) – 3 key criteria**
  - Can we be confident in the exposure characterization?
  - Can we be confident in the outcome assessment?
  - Does the study design or analysis account for important confounding and modifying variables?
- **Experimental Studies (most animal) – 1 key criteria**
  - Can we be confident in the outcome assessment?

Questions?